Précis and Replies to Contributors for book symposium on *Accuracy and the Laws of Credence* to be published in *Episteme*

Richard Pettigrew
Richard.Pettigrew@bris.ac.uk

October 29, 2016

1 Précis of *Accuracy and the Laws of Credence*

I know that all echidnas are monotremes. So I should be at least as confident that all echidnas lay eggs as I am that all monotremes lay eggs — the latter proposition, after all, entails the former. But why must my credences be related in this way? Why must I always be at least as confident in a proposition as I am in any proposition that entails it?

One standard answer is given by the so-called Dutch book argument. Credences play a pragmatic role in my life. They guide my actions; in particular, they guide my betting behaviour. The higher my credence in a proposition, the higher the price I should consider fair for a bet that pays £1 if the proposition is true and £0 if it is false. Thus, if I have greater credence that all monotremes lay eggs than I have that all echidnas do, then I will be prepared to buy a £1 bet on the former at a higher price than I'll be prepared to sell a £1 bet on the latter. And it is straightforward to see that, if I do that, I am guaranteed to lose money overall once the bets pay out. That is, if I am more confident in one proposition than I am in another than it entails, my credences sanction betting behaviour that will lead to a sure loss for me; they leave me vulnerable to a so-called *Dutch book*. I do not deny that Dutch book arguments have some force; but they seem to miss something of the badness of violating the laws of credence that they seek to justify.

Our credences, like our other doxastic states such as full belief, play (at least) two roles: they guide our actions; and they attempt to represent the world. Some think that the action-guiding role is primary, while the world-representing role is secondary: they might think, for instance, that what it means to represent the world accurately is to guide actions in a way that gets us what we want. For them, Dutch book arguments might be completely satisfactory. But others think that the representational role of credences is primary, and the action-guiding role is secondary: credences should guide our actions precisely because their primary aim is to represent the world, and credences that do this better tend to guide use to actions that better get us what we want. Still others think neither is primary and neither secondary — they are simply separate and incomparable roles. For the latter two sets of people, Dutch book arguments will not provide wholly satisfactory argument for the laws of credence, for they say nothing about why credences that violate those norms play their representational role poorly. *Accuracy and the Laws of Credence* attempts to make good this lack. In the book, we consider a suite of putative laws of credence: Probabilism, which says that an agent's credences at a time should satisfy the laws of probability; various version of the Principal Principle, which says how credences about the objective chances should cohere with credences about other propositions; the Principle of Indifference and its variants, which concern the appropriate way to distribute credence in the absence of evidence; and finally, a synchronic version of Conditionalization, which tells us how we should *plan* to update our credences upon receipt of new evidence. The first and fourth comprise standard Bayesian epistemology; the second is a crucial tenet of statistical inference; and the third is the distinctive claim of objective Bayesianism. In each case, we see how the law is supported by considering the representational role that the credences it governs are required to play — credences that violate it will be in some way suboptimal in their attempt to play that role.

We begin with an account of how well a given credal state represents the world; or, as we will say, how accurate it is. The idea is this: for any given possible world and any set of propositions that our agent entertains, there is a set of credences in those propositions that represents that possible world perfectly; a set that is maximally accurate. It is the set that assigns maximal credence of 1 to all truths and minimal credence of 0 to all falsehoods. Let's call that the *perfect* or *ideal credal state* for that possible world. Then we say that a credal state is more accurate at a possible world — and thus better at representing that world — the closer it lies to the credence function that is perfect or ideal at that world. But how do we measure how far one credal state lies from another? In chapter 3, I consider existing accounts of this measure, and find them wanting. In chapter 4, I offer a characterization of the one true measure, the so-called *squared Euclidean distance* (SED). That is, I lay down axioms and show that only one such measure satisfies them all, and it is SED. Putting these components together, we have an account of accuracy that is mathematically precise; we have a precise measure of how well a credal state plays its representational role. The *in*accuracy — and thus epistemic *dis*value — of credal state at a world is its squared Euclidean distance from the ideal credal state. Its accuracy (and thus epistemic value) at that world is just the negative of its inaccuracy (and epistemic disvalue). We formulate this precise account of accuracy in Part I of *Accuracy and the Laws of Credence*. In the remainder, we put it to work in our arguments for the putative laws of credence that concern us in the book.

We begin, still in Part I, arguing in favour of Probabilism. Suppose I offer you the following two options: on the first, I will give you £10 if it rains tomorrow, and you will give me £5 if it doesn't; on the second I will give you £5 if it rains tomorrow, and you will give me £10 if it doesn't. Surely you would be irrational were you to choose the second option. The reason is that, however the world turns out, whether it rains or not, the outcome of the first option is better for you than the outcome of the second — in this situation, we say that the former option *strictly dominates* the latter. For the same reason, it turns out, it is irrational to violate Probabilism. For it turns out that, for any set of credences that violates Probabilism, there is an alternative set on the same set of propositions that is more accurate however the world turns out; that is, it is guaranteed to be better; it is knowable *a priori* that it is better. Thus, the agent who violates Probabilism is in the same position as you are if you pick the second of the two options I offer you above. In both cases, the option picked is strictly dominated. Thus, we can see why any credences that violate Probabilism play the representational role of credences suboptimally — there are alternative credences that are sure to play it better! Thus, this accuracy-based argument for Probabilism complements the Dutch book argument for that law of credence: where the Dutch book argument shows that an agent who violates

Probabilism has credences that guide action suboptimally, because they sanction betting behaviour that leads to a sure loss, the accuracy-based argument for Probabilism shows that an agent who violates that law has credences that represent the world suboptimally, because there are others that are sure to do it better. The mathematical fact on which the accuracy-based argument turns is due to Bruno de Finetti (1974); its philosophical interpretation as an a purely epistemic argument is due to James M. Joyce (1998). In the first part of the book, I address some issues with existing versions of this argument and provide what I take to be its strongest version.

In Part II of the book, we turn to the chance-credence principles, chief amongst them the Principal Principle. These laws say how our credences in propositions about the objective chances should cohere with our credences in other propositions. In this part of the book, we are in fact as occupied with formulating the relevant principles as with justifying them. As David Lewis (1980, 1994) observed, problems arise for intuitive chance-credence principles when the objective chances are *self-undermining*. On many accounts of the metaphysics of objective chance — including the Humean account that Lewis himself favoured — the objective chances themselves assign probabilities to the various hypotheses concerning which probability function gives the objective chances; what we might call the *chance hypotheses*. Thus, in particular, each possible objective chance function assigns a probability to the chance hypothesis that says it gives the objective chances. What's more, for some metaphysics of chance — again, including Lewis' Humeanism — some of them sometimes assign to themselves a probability of less than 1. In this situation, we say that the objective chances are *self-undermining*. And when there are self-undermining chances in play, various natural formulations of chance-credence norms have unpalatable consequences, as we will see below.

When there are no self-undermining chances, on the other hand, we can formulate the Principal Principle in many different ways that turn out to be equivalent: my (conditional) credence in a proposition conditional on its chance being r should be r; my (unconditional) credence in a proposition should be my expectation of its objective chance; my (unconditional) credence in a proposition should be some weighted average of its possible chances. In this situation, we give the following accuracy-based argument in favour of this law of credence. Suppose I offer you two new options: on the first, I give you £10 if it rains tomorrow and you give me £5 if it doesn't; on the second, you give me £10 if it rains tomorrow and I give you £5 if it doesn't. Neither option strictly dominates the other. But suppose further that you know that the objective chance of rain tomorrow is at least 60%. You don't know anything more than that, but you do know that. As a result, while you do not know the actual objective expected payout of the two options, you do know how they are ordered — you know that the objective expected payout of the first is higher than that of the second. Whether the objective chance of rain tomorrow is 60% or something higher, the first option has greater objective expected value than the second. In such a situation, we say that the first option *strictly chance dominates* the second. And we surely judge that if you were to take that second, strictly chance dominated option, you would be irrational. Now, as it turns out, someone who violates the Principal Principle is irrational in exactly the same way. There is an alternative set of credences defined on the same set of propositions that each and every possible objective chance assignment expects to have greater accuracy. That is, there are alternative credences that are guaranteed to have greater objective expected accuracy — they strictly chance dominate the credences that violate the Principal Principle. Every possible chance assignment expects them to better play the representational role of credences. This is our argument for the Principal Principle in chapter .

Now, as noted above, if there are self-undermining chances in play, things don't go so smoothly. Suppose we demand, as we do in the first formulation of the Principal Principle above, and as Lewis (1980) hoped to do, that an agent's conditional credence in a proposition conditional on the claim that the chance of that proposition is r is r. Then this forces the agent to assign no credence whatsoever to any chance hypothesis on which the chances are self-undermining. Or suppose we demand, as in the second formulation above, and as Jenann Ismael (2008, 2013) proposed we do, that an agent's credence in a proposition is her expectation of the objective chance of that proposition. Then we force the agent to assign no credence whatsoever to any chance hypothesis on which the chances are self-undermining and assign non-zero probability to an alternative chance hypothesis on which the chances are not self-undermining. So two out of the three formulations of the Principal Principle given above have undesirable consequences — they constrain an agent too tightly. How are we then to tell what to do in this situation?

So far in *Accuracy and the Laws of Credence*, we have seen the first of two appealing features of the accuracy framework: it can provide non-pragmatic arguments in favour of laws of credence; and these arguments complement the existing pragmatic arguments provided by Dutch book considerations. But at this point, having seen the failure of two putative chance-credence principles in the presence of self-undermining chances, we begin to see its second appealing feature. There are many situations for which we don't know the laws of credence that govern them. In these situations, we don't yet have the putative laws of credence that we wish to justify. But it turns out that the accuracy framework can not only *justify* laws of credence that we have already formulated; it also gives a systematic way to *discover* new laws of credence that we have yet to formulate. We simply ask what properties our credences must have in order to promote the goal of accuracy. We've seen that they must be probabilistic in order to promote that goal; if they are not there will be others guaranteed to better represent the world. And we've seen that, at least when there are no self-undermining chances in play, they must satisfy the Principal Principle; for if they don't there will be others guaranteed to be expected by the objective chances to better represent the world. Now we have a situation, one in which self-undermining chances are in play, where the two laws we formulate ourselves have unwanted consequences. In this situation, we turn to the accuracy approach and ask what it recommends. It turns out that it is this: your credences should be some weighted average of the possible chances. If they are not, there are others that are guaranteed to have greater objective expected accuracy. In the absence of self-undermining chances, this principle is equivalent to the other two formulations of the Principal Principle. In their presence, it is not — it is strictly weaker than both, and does not have their undesirable consequences. So the accuracy framework gives us the chance-credence principle that we seek, and justifies it at the same time.

Part III is devoted to epistemic risk. Suppose I offer you two new options: the first pays you £10,000,000 if a fair coin lands head and £0 if it lands tails; the second pays you £1,000 however it lands. While it's clear that the first has a vastly greater objective expected payoff, it is also the riskier option — the second gives you a guaranteed amount of money; on the first, you risk receiving nothing. Thus, if you are extremely risk averse, you might choose the former option over the latter. What does the accuracy framework say if you take a similar attitude to your credences? What does it recommend if you wish to minimize your epistemic risk as much as possible? In that situation, you want credences whose worst-case accuracy — the accuracy they have when they have least accuracy — is greatest. And it turns out that the credences that do that are those demanded by the Principle of Indifference — if we spread our credences equally over all possibilities, we minimise how inaccurate our credences can possibly be. Thus, the Principle of Indifference is a consequence of extreme epistemic risk aversion or conservatism. In the remainder of this part of the book, we explore the consequences of other attitudes to risk, such as extreme risk-seeking, extreme regret aversion, and the sort of compromises between extreme risk-aversion and extreme risk-seeking that are modelled by the Hurwicz criterion.

Finally, in Part IV, we ask how we should incorporate new evidence into our credences; more precisely, we ask how we should *plan* to incorporate any new evidence. For I argue in chapter 16 that there can be no truly diachronic law of updating — while we might plan to update our credences in a particular way, it is not irrational to abandon that plan and rethink our credences completely when new evidence comes in. In chapter 15, however, we provide three accuracy-based arguments for the synchronic version of the standard Bayesian law of Conditionalization, which says that an agent should plan to update her credences by conditionalizing on the strongest proposition she learns as evidence. The first is due to Hilary Greaves and David Wallace. They argue that Conditionalization maximizes expected accuracy from the point of view of the agent's own credences. But what can this mean? We have been talking so far only of the accuracy of credences, and how well credences represent the world. But we have said nothing about the accuracy of an updating plan. In fact, as Greaves and Wallace note, it turns out to be straightforward to extend our account of the accuracy of credences so that it covers updating rules as well: the accuracy of an updating rule at a given possible world is just the accuracy of the credences that rule would sanction were you receive whatever evidence it is that you would receive at that world. Understood in this way, the rule of Conditionalization is expected to be more accurate than any alternative rule is expected to be (Greaves & Wallace, 2006). The second argument for the synchronic version of Conditionalization turns on a different fact: if you plan to update by any rule that isn't Conditionalization, then there are alternative credences that all of your planned future credences expect to be more accurate than they expect your current credences to be; that is, if you follow through on your plan, you are guaranteed to think of your prior credences as having been suboptimal even at the time they were held. The final argument is due to joint work with Rachael Briggs. In this argument, we don't assess updating plans from the point of view of our prior credences (as in the first argument), or our prior credences from the point of view of our planned future credences (as in the second argument); rather, we assess our priors and plan together as a package. We take the accuracy of such a package to be the sum of the accuracies of its components, the prior and the plan. We show that, if your plan isn't to conditionalize on your new evidence, then there is an alternative package — an alternative prior and plan — that is guaranteed to have greater combined accuracy than your package has (Briggs & Pettigrew, ms).

This completes our brief tour of the central arguments of *Accuracy and the Laws of Credence*. By the end, we hope to have shown why Probabilism, the Principal Principle, and (the synchronic version of) Conditionalization are required by accuracy considerations — that is, we hope to have provided epistemic complements to the pragmatic considerations in favour of these principles that Dutch book arguments provide. We hope to have shown that agents who violate these laws will thereby have credences that will play the representational role of credences suboptimally. And we hope to have shown how we might rationalize the Principle of Indifference — it is connected to the agent's attitude to risk in the epistemic context. What's more, we hope to have developed a new tool that can be used to craft laws of credence governing situations in which it is not clear what we should do. The accuracy framework is already being used in this way to develop laws of credence in the presence of self-locating credences (Kierland & Monton, 1999), peer disagreement (Moss, 2011; Staffel, 2015; Levinstein, 2015), and higher-order evidence (Schoenfield, ta), to name only a few. I hope that *Accuracy and the Laws of Credence* makes a compelling case for this approach, illustrates it strengths, and provides it with a firm foundation.

2 Reply to Levinstein

Link to Levinstein contribution

As noted in the précis, each epistemic utility argument in *Accuracy and the Laws of Credence* has two substantial premises: the first identifies the legitimate measures of epistemic value; the second is a law of rational choice. The first premise remains unchanged in each of the epistemic utility arguments, while the second changes in each. Levinstein's fascinating and searching critique, 'Accuracy Uncomposed', focuses on my argument for the shared first premise.

The argument roughly runs as follows (Chapter 4): the sole fundamental source of epistemic disvalue is inaccuracy (VERITISM); the inaccuracy of a credence function is its distance from the credence function that gets everything right (PERFECTIONISM); the credence function that gets everything right assigns maximal credence to all truths and minimal credence to all falsehoods, which I call the *omniscient credence function* (ALETHIC VINDICATION); and legitimate measures of the distance of one credence function from another have four properties — each is additive (DIVERGENCE ADDITIVITY), continuous (DIVERGENCE CONTINUITY), symmetric (SYMMETRY), and decomposes in a particular way (DECOMPOSITION). As I prove in Theorems 4.3.3 and 4.4.1, the only measure of distance that satisfies these latter four conditions is squared Euclidean distance. Thus, in combination with the first three axioms, we have that the only legitimate measure of inaccuracy is the Brier score, which suffices for all of the epistemic utility arguments given in the book.

Levinstein focusses his attention on DECOMPOSITION. He doesn't deny this axiom. Rather, he argues that an accuracy-first epistemologist — and, in particular, one who motivates DIVERGENCE ADDITIVITY as I do — cannot appeal to it as an axiom when characterizing the legitimate inaccuracy measures. DECOMPOSITION is something we must derive and explain on the basis of an account of accuracy that doesn't assume it.

In section 4.3 of the book, I motivate DECOMPOSITION by noting two conflicting intuitions. I note that one is stronger than the other, and I note that DECOMPOSITION effects a compromise between them that retains the stronger one completely and the weaker one partially. The first intuition is TRUTH-DIRECTEDNESS, which says this: suppose c and c' are credence functions; then, if c assigns a higher credence to each truth than c' does, and if c assigns a lower credence to each falsehood than c' does, then c is less inaccurate (more accurate) than c'. The second intuition is CALIBRATIONISM. Suppose I have credences in the following 100 propositions: *It will rain 1 day from today*, ..., *It will rain 100 days from today*. And suppose that in fact it will rain on 80 out of the 100 days following today. Then it seems that I get something right — I match the world in some important way — if I assign credence 0.8 to each of the propositions (Ramsey, 1931; Shimony, 1988; van Fraassen, 1983; Lange, 1999). In general, according to this intuition, my credences are ideal — they match the world perfectly — if they are *well calibrated*, that is, if each credence matches the proportion of truths amongst all propositions to which I assign that same credence. Now, for any world, there are many well calibrated credence functions — assigning 0.8 to each of the propositions about rain above is well calibrated, since 80% of them are true, but so is assigning 1 to each of the true propositions and 0 to each of the false ones, since 100% of the former and 0% of the latter are true. So how do we measure inaccuracy when this is our account of perfection? My *calibration inaccuracy* is my distance from my *well calibrated counterpart*, which is the credence function that is (i) well calibrated and (ii) assigns the same credence to two propositions whenever I do — equivalently, it is the credence function that assigns as credence to a proposition the proportion of truths amongst all propositions to which I assign the same credence. This contrasts with my *alethic inaccuracy*, which is my distance from the omniscient credence function.[1]

The problem is that these two principles, TRUTH-DIRECTEDNESS and CALIBRATIONISM, each motivated by intuition, are incompatible. Suppose $c(X) = 0.5$ and $c(\overline{X}) = 0.5$. Then c is guaranteed to be well calibrated: whether X is true or false, 50% of the propositions to which I assign credence 0.5 are true. Thus, its well calibrated counterpart c^w is just c itself. Suppose now that X is true. And let $c_\varepsilon(X) = 0.5 + \varepsilon$ and $c_\varepsilon(\overline{X}) = 0.5 - \varepsilon$ (for $0 < \varepsilon < 0.5$). Then, c_ε assigns a higher credence to each truth than c does; and c_ε assigns a lower credence to each falsehood than c does. So, by TRUTH-DIRECTEDNESS, c_ε is more accurate than c. But, by the lights of CALIBRATIONISM, c_ε is less accurate than c, because while c is well calibrated and thus maximally close to its well calibrated counterpart $c^w = c$, c_ε is not well calibrated and thus not maximally close to its well calibrated counterpart, $c^w_\varepsilon = v_w$. How to resolve this tension? I contend that the intuition in favour of TRUTH-DIRECTEDNESS is stronger than the intuition in favour of CALIBRATIONISM; but the intuition in favour of CALIBRATIONISM is nonetheless powerful. Thus, I propose that we attempt to retain the former completely and the latter to the greatest extent that is possible without straying into inconsistency. As Levinstein points out, we need not abandon our monism about epistemic value by saying that calibration accuracy is another source of epistemic value alongside alethic accuracy. So we should not take CALIBRATIONISM to give an account of a novel form of epistemic value. Rather, we should take the intuition behind it to teach us something about the distance measure used to generate the correct account of epistemic value in the presence of ALETHIC VINDICATION.

What the intuition behind CALIBRATIONISM teaches us is suggested by our example above. There c has greater calibration accuracy c_ε, but lower alethic accuracy, and the alethic accuracy of its well calibrated counterpart $c^w = c$ is quite low, for $c^w = c$ lies quite far from the omniscient credence function v_w. On the other hand, c_ε has lower calibration accuracy than c, but higher alethic accuracy, and the alethic accuracy of its well calibrated counterpart $c^w_\varepsilon = v_w$ is maximally high, for $c^w_\varepsilon = v_w$ is maximally close to the omniscient credence function v_w. Thus, while c is closer to c^w than c_ε is to c^w_ε, c^w_ε is closer to v_w than c^w is. So CALIBRATIONISM is incompatible with TRUTH-DIRECTEDNESS because it is possible to get closer to your well calibrated counterpart by moving to a credence function with a different well calibrated counterpart that is itself further from omniscience. This suggests that the alethic accuracy of a credence function is some combination of its calibration accuracy and the alethic accuracy of its well calibrated counterpart. This gives us DECOMPOSITION,

[1] Though it's worth noting that, in some cases, my well calibrated counterpart will also be the omniscient credence function, and in these cases, my alethic and calibrationist accuracy coincide.

which says that there are weights α, β such that $\mathfrak{D}(v_w, c) = \alpha\mathfrak{D}(c^w, c) + \beta\mathfrak{D}(v_w, c^w)$. That is, $\mathfrak{I}(c, w) = \alpha\mathfrak{D}(c_w, c) + \beta\mathfrak{I}(c_w, w)$. The crucial consequence of this axiom is that it preserves a *ceteris paribus* version of CALIBRATIONISM. CALIBRATIONISM says that one credence function is better than another if the calibration inaccuracy of the first is less than the calibration accuracy of the second. The *ceteris paribus* version says that one credence function is better than another *with the same well calibrated counterpart* if the calibration inaccuracy of the first is less than the calibration inaccuracy of the second. Thus, as promised, DECOMPOSITION preserves TRUTH-DIRECTEDNESS in its entirety, and CALIBRATIONISM to the greatest extent possible within the bounds of consistency.

Levinstein raises three objections against this motivation for DECOMPOSITION. Let me consider them in turn. The first, raised in section 2.2.1 of his paper, questions the strength of the intuition in favour of CALIBRATIONISM. Recall: I take TRUTH-DIRECTEDNESS to enjoy greater intuitive support than CALIBRATIONISM, but I take CALIBRATIONISM to enjoy significant intuitive support all the same. Levinstein claims that CALIBRATIONISM only enjoys significant intuitive support in certain cases, such as the example given above of the agent with credences in the 100 propositions about rain. But CALIBRATIONISM is a general claim about accuracy that covers *any* credence function of *any* agent, regardless of the propositions to which she assigns credences. Levinstein gives three examples — c_1, c_2, c_3 — of cases in which we don't have a strong intuition in favour of CALIBRATIONISM. He concludes that the full version of CALIBRATIONISM enjoys very weak intuitive support.

I agree that we lack intuitions in favour of CALIBRATIONISM in the cases Levinstein describes. But I submit that this does not weaken the intuitive case for CALIBRATIONISM and thus for the *ceteris paribus* version that we seek to preserve in DECOMPOSITION. After all, in the presence of other axioms in our characterization of the legitimate inaccuracy measures, it turns out that a strong intuition in favour of a conclusion in just one type of case is sufficient to justify that conclusion in full generality. From PERFECTIONISM, ALETHIC VINDICATION, and DIVERGENCE ADDITIVITY, we know that the inaccuracy of a credence function is the sum, over each of the credences that it assigns, of the distance that credence lies from the corresponding omniscient credence. But note what that entails: Regardless of the credence function and regardless of the opinion set over which it is defined, it is the same distance function that gives the distances from each credence it assigns to the corresponding omniscient credence, and which distances are then summed to give the inaccuracy of the whole credence function. Thus, we can fix general facts about how the inaccuracy of *any* credence function over any opinion set is measured by fixing facts about how the inaccuracy of *some smaller range* of credence functions over a smaller range of opinion sets is measured. Here's an analogy: A codebreaker sits with thousands of coded messages before her; she knows that they were all coded using the same cipher; and she has uncoded versions of some subset of them; then it might be that the uncoded versions of just the subset of them reveal sufficient facts about the cipher to pin it down exactly; and this thereby allows her to decode the other messages. So suppose we have a range of credence functions over a range of opinion sets, and a constraint on legitimate measures of the inaccuracy of *just those* credence functions (for instance, the intuition in favour of CALIBRATIONISM); and suppose that, by placing the constraint on the legitimate measures of inaccuracy *only for those credence functions* we can derive, in conjunction with PERFECTIONISM, ALETHIC VINDICATION, and DIVERGENCE ADDITIVITY, some particular property Φ of the legitimate measures of the distance from one credence to another, which distances from credence to corresponding omniscient credence are then summed to give the legitimate measures of inaccuracy for those credence functions.

Then we can infer that the inaccuracy of *every* credence function is given as the sum of the distances from the credences it assigns to the corresponding omniscient credence, where that distance has the property Φ that we derived. Thus, to support CALIBRATIONISM, and thus in turn the *ceteris paribus* version that is preserved in DECOMPOSITION, we need only have good intuitive support for it in a certain range of cases. And it turns out that the range of credence functions for which Levinstein admits the strength of the intuition in favour of CALIBRATIONISM — that is, the range of cases in which the weather forecaster assigns probabilities to rain in the next n days, for reasonably large n — is sufficiently rich to do precisely this.

Now, while I think this answers Levinstein's objection as he states it, you might feel that he underplays the power of the examples that he gives. He only claims that, in those examples, we *lack* the intuition *in favour* of CALIBRATIONISM. But you might think that, moreover, we *have* an intuition *against* CALIBRATIONISM in at least some of the cases he describes. In fact, I agree. But I do not consider that a mark against my argument for DECOMPOSITION. After all, my argument for DECOMPOSITION explicitly acknowledges that there are cases in which, intuitively, CALIBRATIONISM gives the wrong answer — the case of c and c_ε from above is such a case! I do not motivate DECOMPOSITION by arguing that CALIBRATIONISM is intuitive in all cases — in the comparison of c and c_ε, it clearly isn't. Rather, I argue that it is intuitively supported in sufficiently many cases that we should preserve what we can of it. So Levinstein's examples could only create a problem for this argument if they were to tell not only against CALIBRATIONISM, but against the axiom I posit as containing the kernel of truth in CALIBRATIONISM, namely, DECOMPOSITION. But, in fact, DECOMPOSITION exactly predicts our intuitive reactions to Levinstein's second and third examples, at least in the presence of certain other axioms we assume. DECOMPOSITION (together with DIVERGENCE CONTINUITY) agrees that a credence function, c_2, that assigns 100 slightly different credences in the neighbourhood of 0.6 to 100 different propositions is reasonably accurate if 60 of those propositions are true, and indeed pretty similar in its accuracy to the credence function, $c_{2'}$, that assigns exactly 0.6 to each of the 100 propositions. This is because: (i) $c_2^w = v_w$, so c_2^w has maximal alethic accuracy, so by DECOMPOSITION the alethic accuracy of c_2 is just its calibration accuracy, i.e., the distance from $c_2^w = v_w$ to c_2; (ii) $c_{2'} = c_{2'}^w$, so $c_{2'}$ has maximal calibration accuracy, so by DECOMPOSITION the alethic accuracy of $c_{2'}$ is just the alethic accuracy of $c_{2'}^w = c_{2'}$, i.e., the distance from v_w to $c_{2'}$; and (iii) since c_2 and $c_{2'}$ are close together, the distance from v_w to c_2 is similar to the distance from v_w to $c_{2'}$. DECOMPOSITION (together with DIVERGENCE ADDITIVITY and DIVERGENCE CONTINUITY) also agrees that if 80 out of 100 propositions are true, and 60 out of a different 100 propositions are true, then a credence function $c_{3'}$ that assigns credence 0.8 to each of the propositions in the first set and 0.6 to each of the propositions in the second set is more accurate than a credence function c_3 that assigns credence 0.7 to each of the 200 propositions, even though both c_3 and $c_{3'}$ are well calibrated. This is because those three principle characterize the additive and continuous strictly proper inaccuracy measures; and it is straightforward to prove that, for any two well calibrated credence functions, if one assigns the same credence to two propositions whenever the other does, but not vice versa (as is the case with c_3 and $c_{3'}$) then the second is more accurate than the first.

Levinstein's second objection, raised in section 2.2.2 of the paper, concerns a clash between my argument for DECOMPOSITION and one of my earlier arguments for DIVERGENCE ADDITIVITY. The latter argument turns on the claim that, since an agent's credence function is simply an agglomeration of their individual credences in individual propositions, the accuracy of the credence function should be a function of the individual accuracies of those individual credences in the individual propositions — it should not depend on the 'shape' of the credence function, or other global features of it. Call that LOCALISM. However, as Levinstein points out, only pages later in my argument for DECOMPOSITION, I seem to abandon LOCALISM when I say that the calibration accuracy of a credence function, which very much does depend on its 'shape', is an important component of its overall (alethic) accuracy. What's going on? Well, my response in this case is close to my response to the first objection above. We have an intuition in favour of LOCALISM, and an intuition in favour of CALIBRATIONISM. As Levinstein points out, these are incompatible. But so are CALIBRATIONISM and TRUTH-DIRECTEDNESS, and the intuition in favour of TRUTH-DIRECTEDNESS is the stronger. So we already know that CALIBRATIONISM has to go. What we would like to do, if possible, is to preserve as much of that intuition as is compatible with the other intuitions we cite. We introduce DECOMPOSITION in an attempt to reconcile TRUTH-DIRECTEDNESS and CALIBRATIONISM, but it also reconciles DIVERGENCE ADDITIVITY and CALIBRATIONISM. By showing that there are inaccuracy measures that satisfy DIVERGENCE ADDITIVITY and DECOMPOSITION, we see that it is possible to preserve the intuition for LOCALISM that lay behind our argument for the former axiom, whilst also preserving what is right about the intuition for CALIBRATIONISM that lay behind our argument for the latter.

Levinstein's third objection, raised in section 2.2.3 of his paper, expands on his general contention that DECOMPOSITION, whilst plausible, is a fact that the veritist must derive from an account of accuracy that does not assume it explicitly; it is not something that the veritist can assume. Levinstein's objection is that we should be wary of our intuition in favour of CALIBRATIONISM, the intuition that plays a crucial role in our argument for DECOMPOSITION, because it conflicts so badly with our intuition in favour of TRUTH-DIRECTEDNESS. My response to this clash is to preserve TRUTH-DIRECTEDNESS completely and CALIBRATIONISM only partially. But Levinstein suggests that we should be less forgiving of CALIBRATIONISM — its clash with TRUTH-DIRECTEDNESS drains it of all force, at least for a veritist, as I claim to be. It is important, I think, that the objection is directed particularly against a veritist who appeals to the intuition in favour of CALIBRATIONISM, and not against just anyone trying to place constraints on a measure of epistemic utility. For it is certainly not in general bad methodology in normative theorising to appeal to two conflicting intuitions and try to save one completely and the other in part. For instance, in ethics, when I formulate my account of moral actions, I might appeal to rights-based intuitions and utilitarian intuitions, notice that they clash in various cases, and use that observation to motivate a lexicographic account on which the morally best action is the one, amongst those that respect all rights, that maximises total utility. This account saves the rights-based intuitions completely and the utilitarian intuitions partially. In fact, it gives a *ceteris paribus* version of the utilitarian intuition that is analogous to the *ceteris paribus* version of CALIBRATIONISM entailed by DECOMPOSITION. But it would not be appropriate to object to this justification of the lexicographic account by claiming that the utilitarian intuition is drained of all its plausibility by its clash with the rights-based intuition — the fully general utilitarian intuition is defeated by that clash, perhaps, but it is worth retaining as strong a restricted version of it as we can. However, such an objection might get some force if I also held a position in metaethics, analogous to veritism in epistemology, that says that the sole fundamental source of moral facts is the domain of rights. Then we might question whether an appeal to the utilitarian intuition is appropriate. And this is analogous to Levinstein's objection. The question is whether it is appropriate for a veritist to appeal to the intuition in favour of CALIBRATIONISM. I think it is. What the clash between TRUTH-DIRECTEDNESS and CALIBRATIONISM reveals is not that the intuition in favour of CALIBRATIONISM is in conflict with veritism. Rather, it reveals that we took that intuition to be more general that it really is when we formulated CALIBRATIONISM. That is, what it reveals is that the intuition is really an intuition in favour of the *ceteris paribus* version of CALIBRATIONISM preserved in DECOMPOSITION.

Levinstein's challenging objections have pushed me to clarify the structure of my arguments for DIVERGENCE ADDITIVITY and DECOMPOSITION. Those arguments are based on intuitions: the intuitions in favour of TRUTH-DIRECTEDNESS, LOCALISM, and CALIBRATIONISM. We note that the latter intuition clashes with the former two; we hold that the former two intuitions are the stronger, even individually and certainly in conjunction. We contrive to preserve the former two intuitions and abandon the latter; but we wish to retain as much of the latter as is possible. That leads us to posit DIVERGENCE ADDITIVITY and DECOMPOSITION.

3 Reply to Staffel

Link to Staffel contribution

In *Accuracy and the Laws of Credence*, my epistemic utility argument for Probabilism — adapted from Jim Joyce's original (Joyce, 1998) — relies on a claim about the value of having a set of credences, and a mathematical fact. The value claim: the value of having credences is given by their purely epistemic value, which is in turn given by their accuracy, which is measured by the negative of the Brier score. The mathematical fact: for each set of credences that violates the probability axioms, there is an alternative set of credences that satisfies those axioms that is guaranteed to be more Brier-accurate; and, for any set of credences that satisfies the probability axioms, those credences expect themselves to be most Brier-accurate. Here's a natural objection to it. The argument claims to establish a law of credence, so it might seem legitimate that we take the value of having a particular set of credences to be just the purely epistemic value of those credences themselves. But in fact it's not, because having credences inevitably leads (or should lead) to other things, alongside the credences, that are evaluable and whose value should be combined with the purely epistemic value of the credences to give the overall value of having those credences. For instance, having a high credence that Trump might win the upcoming US presidential election might lead me to have various emotional responses — I might feel fear, despair, and anxiety. And it might lead me to do certain things — I might contact my American expat friends and encourage them to send in an overseas vote. So the value of having that credence should be given partly by its accuracy, partly by the utility I assign to feeling afraid, despairing, and anxious, and partly by the utility of the act of getting out the expat vote. When these various sources of value are all combined to give the overall value of having the credences, the theorem that underpins the accuracy argument for Probabilism no longer entails that, for each non-probabilistic set of credences, there is an alternative probabilistic set that is guaranteed to have greater value. So the argument fails.

Stated like this, the objection is easily answered. What we seek in *Accuracy and the Laws of Credence* are *purely epistemic arguments* for the laws of credence. Thus, while having a particular set of credences might have other sorts of value — aesthetic, moral, political, subjective, pragmatic — we can ignore these in the measures of value we use in our arguments for the laws of credence because those arguments claim only that, *from the purely epistemic point of view*, an agent should obey these laws. For instance, we do not claim that, if by obeying Probabilism I will bring pain and destitution on my family, I should nonetheless obey that law.

However, there is a version of this objection that avoids this response. It points out that, as well as emotional states and practical decisions, there is another sort of thing to which credences give rise, namely, other doxastic states, such as the categorical doxastic states known to philosophers as *full* or *outright beliefs*. For instance, according to one sort of Lockean view, having a sufficiently high credence in a proposition will always lead me to have a full belief in that proposition; according to another, it should lead me to have that attitude, whether it always does or not. Now, we cannot ignore these putative effects of having a set of credences in the way we dismissed the emotional and decision-making effects above. Those effects were not evaluable for their epistemic goodness. But full beliefs clearly are so evaluable. Thus, when we evaluate the goodness of having a set of credences, we should include not only the epistemic goodness of those credences, but also the epistemic goodness of the full beliefs to which they do (or should) give rise. However, as I explained in section 4.2 of the book (and, in more detail, in (Pettigrew, 2015)), when we combine these two sources of epistemic goodness in the most natural way, the accuracy argument for Probabilism fails. I conclude that, in order to appeal to epistemic utility arguments for the laws of credence, we must jettison full beliefs and posit only credences. That is, I advocate a *doxastic attitude monism* of the sort proposed by Jeffrey (1970); or, what Staffel calls *linguistic reductivism*, since I think that *talk* of full beliefs still makes sense, but it reduces quickly and unproblematically to talk of credences. Staffel argues that my conclusion is too quick. She proposes an account of *doxastic attitude dualism* — the view on which there are both credences and full beliefs — and argues that it is compatible with the epistemic utility arguments given in *Accuracy and the Laws of Credence*. I'm persuaded! And I'm extremely grateful to Staffel for broadening the appeal of the epistemic utility arguments proposed in the book.

According to Staffel, an agent has both credences and beliefs, and in a particular context of reasoning she chooses which combination of these attitudes to use, always ensuring that she never uses both a credence and a belief in the same proposition. Thus, when I try to decide whether or not to take an umbrella when I go outside, I might appeal to my *belief* that there isn't a vindictive umbrella-hater on the loose throwing eggs at all those who carry such an accessory. Doing so allows me to ignore that possibility in my reasoning, and doing that makes my reasoning much more efficient and only very slightly less reliable. But I will use my *credence* that it will rain and my *credence* that it won't in my reasoning. And I'll use them to calculate my expected utility for the action of taking the umbrella, and my expected utility for the action of not doing so. By ensuring that I use only one attitude per proposition in any given context of reasoning, Staffel ensures that, at least when we evaluate the accuracy of the doxastic states that comprise that context of reasoning, we will not encounter the clash between dualism and the epistemic utility arguments that I describe in the book. That clash arises precisely because I took the agent to have a belief in and assign a credence to the same proposition. But perhaps other problems arise when we adopt Staffel's picture? To convince ourselves they don't, let's see how we might spell out her account in more detail.

Here, I will consider two ways of spelling out Staffel's suggestion — she alludes to both herself and suggests that these are where we might look. The first is inspired by Jonathan Weisberg's doxastic attitude dualism (Weisberg, ta); the second by Roger Clarke's version of dualism (Clarke, 2013). For Weisberg, there are credences and there are beliefs. Beliefs may be formed as the result of acquiring a certain sort of high credence, but they are not reducible to this. They are used by our minds when we undertake certain sorts of efficient and reliable reasoning, such as the methods of Elimination by Aspects (Tversky, 1972) and the Evidence Accumulation Model (Lee & Cummins, 2004) posited by psychologists. We have both beliefs and credences in many propositions; but, when it comes to a particular context of reasoning, we only ever use our belief in a proposition or our credence in it, never both. This clearly agrees with Staffel's account of what happens in a context of reasoning. Moreover, it makes claims about what is going on in the background: it tells us what existing resources we draw upon when we find ourselves in one of these contexts of reasoning and need to identify which doxastic states to use.

How does this account fare when we assess the doxastic states in question for their accuracy? The background states — those from which we select the credences and beliefs that we use in a particular context of reasoning, and where credences and beliefs are assigned to the same proposition — clearly give rise to the clash with epistemic utility arguments that lead me originally to my own monism. But this won't be a problem in a given context of reasoning, since by hypothesis we don't ever select a credence and a belief in the same proposition. So it might well be that our epistemic utility argument for Probabilism works within the confines of any given context of reasoning. And, if this were so, we might be able to argue that the background credences must be probabilistic, since we would know that credences in a context of reasoning must be probabilistic, and it is the background credences that are put to use within a context of reasoning. So, the question becomes: does our epistemic utility argument for Probabilism go through within a context of reasoning when we adopt Weisberg's model? Unfortunately, I think not. Suppose I am in a context of reasoning in which I will consider four propositions: *Clinton will win* (C), *Sanders will win* (S), and *Clinton or Sanders will win* ($C \vee S$). My background credences are as follows: $c(C) = 0.9$, $c(C \vee S) = 0.95$, $c(S) = 0.05$. And my background beliefs are: a belief in C, a belief in $C \vee S$, and no belief in S. In the context of reasoning I currently inhabit, I use my full belief in C, my full belief in $C \vee S$, and my credence of 0.05 in S. The problem is that the total score of my background credences and beliefs will be the score of a full belief in C, a full belief in $C \vee S$, and credence 0.05 in S. But assuming that each full belief is scored the same, these states will be scored exactly as would a credence function that assigned the same credence to C and to $C \vee S$, and a credence of 0.05 to S. But of course any such credence function would be incoherent and thus accuracy dominated (provided my scoring rule is strictly proper). So there would be credences I could assign to C, $C \vee S$, and S in this context of reasoning that would be guaranteed to be more accurate than my full beliefs in C and $C \vee S$ and my credence in S. This suggests that Weisberg's proposal runs into difficulties when treated within epistemic utility theory. Thus, we have a concrete instance of precisely the problem that Staffel predicts in the final paragraph of section 4 of her article — we need a mechanism that an agent can use efficiently to select credences and beliefs in a way that makes her coherent.

There are a number of responses to this objection that we might give on behalf of this way of spelling out Staffel's proposal. The first says that the example is flawed: if I use my full belief in a proposition (such as C), I should not also use a positive credence in another proposition (such as S) that is inconsistent with it. This, the response claims, is a guiding principle of reasoning in a context — belief in a proposition rules out possibilities that are incompatible with it; so they shouldn't be assigned positive credence. An initial problem with this response is that we need a more general guiding principle than this. The problem above arose because, in our context of reasoning, we used credences in certain propositions (one proposition in this case, namely, S) that, together with Probabilism, would force credences in two further propositions (namely, C and $C \vee S$) to be distinct; but we then forced those further propositions to always have the same accuracy score because we took them both to be believed; this meant that the accuracy profile of the beliefs and credences in the context of reasoning was identical to the accuracy profile of a set of non-probabilistic credences; and that meant that the beliefs and credences were accuracy dominated. So, in order to avoid the issue identified here, we would need the following guiding principle: if, together, the set of credences I use in a context of reasoning is such that, via Probabilism, they would entail that two propositions are assigned different credences, then I cannot use a belief in both of those propositions. The problem with this, however, is that it requires a lot of cognitive processing to calculate whether I have satisfied this guiding principle or not. And we are using full beliefs precisely to avoid the costs of such processing.

A second response to the objection from above says that the objection misses the point. We appeal to full beliefs in a context of reasoning for purely pragmatic reasons. Had we but computational resources enough and time, we would use only credences in any given context of reasoning. But we lack these luxuries, and so we appeal sometimes to full beliefs. But since this is done purely for pragmatic reasons, and not for epistemic reasons, it is not appropriate to score the full beliefs for their epistemic utility at all. It is as if I were to tie a knot in my handkerchief to remind myself to buy a pomegranate, and you were to judge the knot for the accuracy of its representation of a pomegranate. Full beliefs, in this picture, are not employed for their representational features, but for their assistance in making efficient and reliable decisions; so they should not be valued for their representational accuracy, but instead for their practical assistance in making efficient and reliable decisions.

This is an interesting response, because if it works, it is a general response to the accuracy-based objection to dualism, however you fill in the details. It allows that there are beliefs and there are credences. And it even allows that both sorts of attitude can be ascribed to the same proposition, both in the background and in a context of reasoning. But it avoids my objection to such a dualist picture on the grounds that the epistemic value of the whole doxastic state, which includes beliefs and credences, is not the sum of the epistemic value of the beliefs and the epistemic value of the credences, but rather the epistemic value of the credences alone. However, I think it doesn't work. The reason is that mental states that aren't valued for their accuracy aren't really recognisable as beliefs at all. All that we need in order to implement the Elimination by Aspects and Evidence Accumulation Model that Weisberg considers is some sort of cognitive marker that attaches to a proposition that allows it to be employed in a context of reasoning as if it were true. But this cognitive marker may not be the belief marker. Of course, the argument that it is the belief marker is a functionalist one: it seems to have many of the features that we typically ascribe to beliefs. But if we deprive it of one of the most crucial such features — if we no longer value it for its accuracy or fit to world — then we lose much of the reason to call it a belief. It is no longer a representational state, but purely a mental state that quickens reasoning. And that is not a belief state. So there are, I think, serious problems with Weisberg's picture from an accuracy-first or epistemic utility perspective.

Let me turn next to another of Staffel's suggestions for how to spell out her proposal. It is inspired by Roger Clarke's version of dualism (Clarke, 2013). According to Clarke, we have what he calls a *global credence function*. This acts a little like the background credences in Weisberg's model. It supplies the credences we use in a particular context, which Clarke calls *local credences*. According to Clarke, our local credences in a context are obtained from our global ones by conditionalizing on the proposition that characterizes that context. For Clarke, contexts are sets of possibilities, and so the proposition that characterizes a context is the proposition that is true at all and only the possibilities that comprise it. Thus, I obtain the credences I use at a context by taking my global credences, ruling out all the possibilities that I ignored in that context, and renormalizing. Clarke then says that an agent has a belief in a proposition in a context if, and only if, their local credence in that proposition in that context is 1.

Now, the first thing to note about Clarke's account is that it looks like a version of linguistic eliminativism. Why posit beliefs and credences when you take beliefs to amount to no more than having maximal credence? But suppose we do. Then Clarke's proposal still doesn't quite fit with Staffel's framework, since when I have local credence 1 in a proposition, I *also* have a belief in it, and so the accuracy-based objection threatens. In order to fit Staffel's framework, we must say that our local credences are given by conditionalizing on the context, but we then drop our assignment of a credence to a proposition when that credence is 1, and replace it with a full belief in that proposition.

How does this proposal fare when we assess the doxastic states it posits for their accuracy? Initially, it might seem to run into two problems. First, it might seem to clash with the Principal Principle, for which there is an epistemic utility argument, which I describe in Part II of *Accuracy and the Laws of Credence* (cf. also (Pettigrew, 2013; Caie, 2015)); second, it might seem to clash with Bayesian Conditionalization, for which there are three epistemic utility arguments, which I describe in Part IV of *Accuracy and the Laws of Credence* (cf. also (Greaves & Wallace, 2006; Briggs & Pettigrew, ms)). Let's consider the Principal Principle first. Suppose I know that the objective chance that I will be struck by lightning as I walk to work today is one in ten million. When I am deciding what shoes to wear, I discount this possibility. That is, I reason within a context in which it does not occur. So I conditionalize on a proposition that rules out a lightning strike. Then my local credence that I will be struck by lightning is 0; but I know the chance is one in ten million. Thus, I violate the Principal Principle, which says (at least) that my credences should match any known chances.

I think this is too quick. This shows only that Clarke's model clashes with the following hybrid version of the Principal Principle: My *local* credences should match any chances of which my *global* credence function is certain. But, since the Principal Principle is not usually formulated with the local/global credence distinction in mind, there is no reason to think that is the correct formulation. Indeed, it seems to me that the Principal Principle is a principle of coherence: it tells us how our credences in propositions about the chances should relate to our credences in other propositions. Conceived like this, it seems most natural to say that our global credence function should satisfy it, and each of our local credence functions should, but there is obligation for our local and global credences to jointly satisfy the hybrid version for which we caused trouble above. So the Principal Principle causes no problem for Clarke's framework.

Perhaps the more serious objection concerns the compatibility of Clarke's framework with Bayesian Conditionalization. The problem is that, on Clarke's account, we move from one set of local credences to another whenever the context of reasoning changes. If we are in context C at time t and C' at time t', then our local credences at t are given by our global credences conditional on C and local credences at t' are given by our global credences conditional on C'. This would pose no problem if contexts were entirely determined by our evidence, so that a context changes to C if, and only if, one learns C as evidence and nothing stronger. But that's not how they're supposed to work. Various non-evidential features of an agent's situation influence the context in which they reason at a given time, such as the stakes of decisions that they are about to make at that time. So, on Clarke's view, agents will often violate Bayesian Conditionalization. Given that there is an epistemic utility argument for Bayesian Conditionalization, it might seem that this worrisome for the view. But I think not. One thing to say is that, in Part IV of *Accuracy and the Laws of Credence*, I argue that there is no diachronic requirement to *update* by conditionalization; there is only the synchronic requirement to *plan to update* in this way. But that won't buy us much time, since it's clear that I often know ahead of time how my contexts will change, and Clarke must say that I plan to respond to those changes in context by using local credences obtained from global credences by conditionalizing on these contexts in a way that, as we have seen, violates Bayesian Conditionalization. A better response to this objection to Clarke's view is to note that the arguments for planning to update by conditionalization all assume that the full space of possibilities remains unchanged when the agent acquires evidence. They say nothing about those situations in which an agent expands her conceptual space and thereby grains the set of possible worlds more finely, or shrinks her conceptual space and thereby grains the possibilities more coarsely. For instance, Bayesian Conditionalization, in either its diachronic or synchronic form, has nothing to say about a situation in which I learn to distinguish inertial from gravitational mass, or jadeite and nephrite, and thereby expand the set of possibilities I consider. Nor has it anything to say about the situation in which I've learned that Currer Bell is Charlotte Brontë or that Kwame Ture is Stokely Carmichael, and thereby shrink the set of possibilities I consider. Moving from one context to another in Clarke's model is akin to this: certain possibilities are removed and possibly others are introduced. Bayesian Conditionalization has nothing to say about these cases, and so it cannot count against Clarke's model that it seems to violate that law of credence.

In sum, I warmly welcome the alternative response that Staffel offers to my epistemic utility argument against doxastic attitude dualism. As Staffel observes, my argument shows only that credences and beliefs should not be assigned to the same proposition. And that does not entail that there is just one sort of doxastic attitude. Rather, it is possible that there are both beliefs and credences at work in various contexts of reasoning. I also agree that we can learn from Weisberg's model and Clarke's when filling out the details of this proposal: from Weisberg, we take the stricture against having a belief in and assigning a credence to the same proposition; and from Clarke we take the dynamics of local credences and the connection between maximal credence and full belief.

4 Reply to Cariani

Link to Cariani contribution

Part III of *Accuracy and the Laws of Credence* concerns the so-called *chance-credence principles*: David Lewis' Principal Principle (PP) (Lewis, 1980), Ned Hall's and Michael Thau's New Principle (NP) (Hall, 1994, 2004; Thau, 1994), and Jenann Ismael's General Recipe (GR) (Ismael, 2008), along with variations on that theme, including the principle that I favour in the book, a version of the Evidential Temporal Principle (ETP^+). Roughly speaking, chance-credence principles say how an agent's attitudes towards propositions about the objective chances should relate to her credences in the propositions to which those objective chances assign probabilities. Cariani's excellent 'Chance, credence and circles' raises two worries about the epistemic utility argument that I offer for ETP$^+$.[2] Let me begin, then, by formulating that particular chance-credence principle and describing the argument I offer in its favour. In the process, I hope to clarify some of its features.

Suppose Maya has only a credence in the proposition, H, that the coin she is holding will land heads on its next toss. And suppose that her current total evidence entails that the chance of H is strictly between 60% and 70% or strictly between 80% and 90% — she knows it's a trick coin, perhaps, but doesn't know which of two trick coin factories produced it; one produces coins with biases strictly between 60% and 70%, the other with biases strictly between 80% and 90%. That is, if we let $\mathcal{EC}$ be the set of probability functions that are epistemically possible current chance functions for Maya — probability function that might, for all her evidence says, give the current chances — then

$$\mathcal{EC} = \{ch : 0.6 < ch(H) < 0.7 \text{ or } 0.8 < ch(H) < 0.9\}$$

Now suppose Maya has credence 0.25 in H. Then we would likely judge her irrational — we would say that she has not responded appropriately to her evidence. And similarly if she had 0.3, 0.96, or indeed any credence below 0.6 or above 0.9. On the other hand, if she had credence 0.65, 0.75, 0.83, or indeed any credence strictly between 0.6 and 0.9, these would seem appropriate responses to her evidence. So it seems that one way to respond reasonably to this evidence is to have a credence that is a weighted average of finitely many epistemically possible chances — in the language of geometry, these are the credences in the *convex hull of* $\mathcal{EC}$ (denoted $\mathcal{EC}^+$), where the convex hull of a set of credence functions is the smallest set that contains it and contains a weighted average of two credence functions whenever it contains those credence functions. Thus, for Maya, $\mathcal{EC}^+ = \{c : 0.5 < c(H) < 0.9\}$. Are there other ways? Well, what if Maya had credence 0.6? Or 0.9? Again, these seem reasonable responses to her evidence. But they aren't weighted averages of finitely many possible chance functions. Nonetheless, they are the limits of an infinite sequence of such weighted averages — in the language of geometry, they are the *limit points* of the convex hull of $\mathcal{EC}$, and we obtain the closure of that convex hull (denoted $\text{cl}(\mathcal{EC}^+)$) by adding all such limit points. Thus, for Maya, $\text{cl}(\mathcal{EC}^+) = \{c : 0.6 \leq c(H) \leq 0.9\}$. According to ETP$^+$, my favoured chance-credence principle, these are the two ways in which an agent can appropriately respond to evidence that bears on hypotheses about the objective chances — her credence function may be a weighted average of finitely many of the epistemically possible chance functions; or it may be the limit of an infinite sequence of such weighted averages. Using the notation from geometry that we have just introduced:

> **ETP$^+$** It is a requirement of rationality that an agent's credence function c is in $\text{cl}(\mathcal{EC}^+)$, where $\mathcal{EC}$ is the set of epistemically possible objective chance functions for her.[3]

[2]In fact, Cariani talks of a slightly different version of the Evidential Temporal Principle, namely, ETP. As I note in the book, ETP and ETP$^+$ are equivalent under certain assumptions. But I focus on ETP$^+$ here, since it is the principle that follows directly from my favoured epistemic utility argument without any further assumptions. To obtain ETP from that epistemic utility argument, we must make the assumptions under which ETP$^+$ entails ETP: we must assume that the agent has positive credences only in a set of chance hypotheses that they take to exhaust the possibilities; and we must assume that each of those chance hypotheses posits non-self-undermining chances.

[3]In *Accuracy and the Laws of Credence*, I state this differently. There, ETP$^+$ says that it is a requirement of rationality that c is in $\text{cl}(\mathcal{EC}_E^+)$, where $\mathcal{EC}_E$ is the set of epistemically possible chance functions conditionalized

When I say that a probability function ch is an epistemically possible objective chance function here, I mean the following: let T_{ch} be the proposition that says that the current objective chance function is ch; and let E be the agent's total evidence; then ch is an epistemically possible objective chance function for that agent if E is compatible with T_{ch}.

This, then, is ETP$^+$, my favoured chance-credence principle. And here is the epistemic utility argument in its favour. It is based on the following general mathematical fact:

Theorem 1 *Suppose $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure, such as the Brier score. And suppose $\mathcal{S}$ is a set of probability functions. Then*

(I) *If c does not lie in $\mathrm{cl}(\mathcal{S}^+)$, there is c^* in $\mathrm{cl}(\mathcal{S}^+)$ such that every probability function in $\mathcal{S}$ expects c^* to be more accurate than it expects c to be.*

That is, $\mathrm{Exp}_{\mathfrak{I}}(c^*|p) < \mathrm{Exp}_{\mathfrak{I}}(c|p)$, for all p in $\mathcal{S}$.

In this situation, we say that c is strongly $\mathcal{S}$-dominated by c^*.

(II) *If c does lie in $\mathrm{cl}(\mathcal{S}^+)$, there is no c^* in $\mathrm{cl}(\mathcal{S}^+)$ such that every probability function in $\mathcal{S}$ expects c^* to be at least as good as it expects c to be.*

That is, $\mathrm{Exp}_{\mathfrak{I}}(c|p) < \mathrm{Exp}_{\mathfrak{I}}(c^*|p)$, for some p in $\mathcal{S}$.

In this situation, we say that no c^* even weakly $\mathcal{S}$-dominates c.

Theorem 1 comprises the third premise of the epistemic utility argument for ETP$^+$. The first premise, shared with the other epistemic utility arguments in the book, is **Brier Alethic Accuracy**, which says that the only legitimate measure of epistemic disvalue is the Brier score. And the second premise, which is the focus of Cariani's concerns, is **Current Chance Dominance** (CCD), which is the following decision-theoretic principle:

Current Chance Dominance Suppose

(i) c is strongly $\mathcal{EC}$-dominated by c^*

(ii) c^* is not even weakly $\mathcal{EC}$-dominated by any c'

(iii) c^* is immodest.

Then c is irrational for an agent for whom $\mathcal{EC}$ is the set of epistemically possible chance functions.

Together, with $\mathcal{S}$ taken to be $\mathcal{EC}$ in Theorem 1, these three premises entail ETP$^+$.

Cariani objects to this argument in two ways: first, he objects that it is circular in a particular way; second, he objects that it establishes too much — ETP$^+$ is not a genuine chance-credence principle. I'll consider the second objection first, since it will help to have our response to that in place when we consider the first objection.

Cariani points out that ETP$^+$ is a narrow scope norm. It says: *if* an agent has a certain body of evidence, *then* she should have a credence function in the closure of the convex

on the agent's total evidence E, $\mathcal{EC}^+_E$ is the convex hull of this set, and $\mathrm{cl}(C^+_E)$ is the closure of that convex hull. Here, I drop the conditionalization on E. The reason is that, since writing the book, Jason Konek has convinced me that conditionalizing the possible chance functions on E is redundant. After all, if I have evidence E now, then E must be true now, and so whatever the actual chance function is now, it must incorporate E by assigning it maximal probability, in which case conditionalizing on E will have no effect. In other words, $\mathcal{EC} = \mathcal{EC}_E$, and so $\mathrm{cl}(\mathcal{EC}^+) = \mathrm{cl}(\mathcal{EC}^+_E)$.

hull of the set of putative objective chance functions that are compatible with that evidence. Chance-credence principles, on the other hand, are wide scope norms — or so Cariani contends. They are principles of coherence; they say how credences in one sort of proposition — namely, chance hypothesis — relate to credences in another sort of proposition — namely, the propositions to which the objective chances assign a probability. They say nothing about how those credences should relate to the agent's evidence. Thus, ETP$^+$ is not a genuine chance-credence principle and the epistemic utility argument in its favour proves too much.

I agree with Cariani. While I maintain that ETP$^+$ is a principle of rationality for credences, it might be better not to advertise it as a pure chance-credence principle. Rather, it is the result of conjoining a wide scope chance-credence principle and a narrow scope evidential principle. We might call the wide scope principle TP$^+$ and formulate it thus:

> **TP$^+$** It is a requirement of rationality that an agent's credence function c is in $\mathrm{cl}(\mathcal{MC}^+)$, where $\mathcal{MC}$ is the set of metaphysically possible objective chance functions.

And we might call the narrow scope principle E and formulate it thus:

> **E** It is a requirement of rationality that an agent's credence function assign maximal credence to her total evidence.

Then we can see that $E + TP^+ = ETP^+$.[4]

Moreover, we can give epistemic utility arguments for E and for TP$^+$. As usual, the first premise of each is Brier Alethic Accuracy. The second premise of the argument for E is a narrow scope strengthening of Dominance: it says that a credence function c is irrational if there is another c^* that is more accurate than c at all *epistemically* possible worlds, and there is no c' that is at least as accurate as c^* at all *epistemically* possible worlds. If we take a world to be epistemically possible if the agent's total evidence is true at that world, then this version of Dominance (together with Brier Alethic Accuracy) entails E. The second premise of the argument for TP$^+$, on the other hand, is a wide scope weakening of Current Chance Dominance: it says that a credence function c is irrational if there is another c^* that

[4] *Proof.*

(i) E + TP$^+$ $\subseteq$ ETP$^+$. Suppose c satisfies E and TP$^+$. There are two options:

  (a) c is in $\mathcal{MC}^+$. That is, c is a weighted average of finitely many metaphysically possible chance functions. Now suppose ch is an epistemically impossible chance function — that is, T_{ch} and E are incompatible. Then, since $c(E) = 1$, $c(T_{ch}) = 0$. And, since $ch(T_{ch}) > 0$, c must give no weight at all to ch. Thus, c is a weighted average only of members of $\mathcal{EC}$. That is, c is in $\mathcal{EC}^+ \subseteq \mathrm{cl}(\mathcal{EC}^+)$. So c satisfies ETP$^+$.

  (b) c is a limit point of $\mathcal{MC}^+$. That is, there is a sequence $c_1, c_2, \ldots$ such that $c_n \to c$ as $n \to \infty$, where each c_n is a weighted average of finitely many members of $\mathcal{MC}$ — that is, each c_n is in $\mathcal{MC}^+$. Thus, $c_n(-|E) \to c(-|E)$ as $n \to \infty$. By a standard result due to (Raiffa, 1968), since each c_n is a weighted average of chance functions in $\mathcal{MC}$, each $c_n(-|E)$ is a weighted average of those chance functions in $\mathcal{MC}$ conditional on E. But each chance function in $\mathcal{MC}$ conditional on E must be a chance function in $\mathcal{EC}$ — if $ch(E) = 1$, then E and T_{ch} must be compatible, because $ch(T_{ch}) > 0$. Thus, each $c_n(-|E)$ is a weighted average of finitely many chance functions in $\mathcal{EC}$ — that is, $c_n(-|E)$ is in $\mathcal{EC}^+$. So $c(-|E)$ is a limit point of $\mathcal{EC}^+$. Thus, $c(-|E)$ is in $\mathrm{cl}(\mathcal{EC}^+)$. By E, $c(E) = 0$, so $c(-|E) = c$. Thus, c is in $\mathrm{cl}(\mathcal{EC}^+)$. So c satisfies ETP$^+$.

(ii) ETP$^+$ $\subseteq$ E + TP$^+$. Suppose c satisfies ETP$^+$. Then c is in $\mathrm{cl}(\mathcal{EC}^+)$. So $c(E) = 1$, since $ch(E) = 1$ for each ch in $\mathcal{EC}$, $ch(E) = 1$. Thus, c satisfies E. And c is in $\mathrm{cl}(\mathcal{MC}^+)$, since $\mathcal{EC} \subseteq \mathcal{MC}$. So c satisfies TP$^+$ as well.

strongly $\mathcal{MC}$-dominates c, and there is no c' that even weakly $\mathcal{MC}$-dominates c^*, and c^* is immodest. That is, it is obtained from Current Chance Dominance by replacing $\mathcal{EC}$ with $\mathcal{MC}$ throughout. This version of Current Chance Dominance (together with Brier Alethic Accuracy) entails TP^+.

Let me turn now to Cariani's first concern about the epistemic utility argument for ETP^+. He worries that it is circular. The argument has three premises: (1) Brier Alethic Accuracy; (2) Current Chance Dominance (CCD); (3) Theorem 1. Where might the circularity lurk? Cariani appeals to Walter Sinnott-Armstrong's helpful distinction between structural and dialectical circularity (Sinnott-Armstrong, 1999). In Section 10.2 of the book, I endeavour to answer the objection that the epistemic utility argument for ETP^+ is structurally circular — that is, its conclusion is explicitly assumed in its premises or the best justification of its premises assumes its conclusion. But Cariani focusses on arguing that it is dialectically circular — that is, it would fail to persuade someone who was not antecedently convinced of the truth of the conclusion.

He considers two ways in which we might test an argument for dialectical circularity. They are based on two reasons an interlocutor not antecedently convinced of the conclusion might reject the argument. Since the argument in question is valid, Cariani considers two reasons why such an interlocutor might reject its premises — in particular, why they might reject CCD.

First, the interlocutor might reject the premise because she judges it 'too close' to the conclusion, even if it doesn't strictly entail it or require it for its justification. Here, the distance of a premise in an argument from its conclusion is presumably measured by the strength of the other assumptions that must be added to that premise to derive the conclusion. So an interlocutor might reject a premise in the argument because she sees that it requires only very weak assumptions to derive from that premise the conclusion that she rejects. Cariani considers this way of raising the dialectical circularity worry against the epistemic utility argument for ETP^+, but thinks it is very hard to judge how close counts as too close, and hard to judge how close CCD is to ETP^+.

I think there is a little more to say here. After all, we know with reasonable precision how much must be added to CCD to obtain ETP^+. We must add at least the assumption that the legitimate inaccuracy measure $\mathfrak{I}$ is strictly proper — whether it need be additive and continuous as well is more complicated. For suppose that $\mathfrak{I}$ is not strictly proper. Then there is a probability function p and a credence function c such that $\text{Exp}_{\mathfrak{I}}(c|p) \leq \text{Exp}_{\mathfrak{I}}(p|p)$. Now, suppose further that p is the only putative current chance function that is epistemically possible for our agent — that is, $\mathcal{EC} = \{p\}$, and so $\text{cl}(\mathcal{EC}^+) = \{p\}$. Then CCD would permit our agent to violate ETP^+, for in this instance, ETP^+ demands of a rational agent that she adopt p as her credence function, while CCD, applied with this improper inaccuracy measure $\mathfrak{I}$ rules that, if p is rational, so is c, and c violates ETP^+.

So we must add to CCD at least the assumption that the legitimate inaccuracy measure is strictly proper. This is a substantial assumption. It rules out the absolute value measure, which takes the inaccuracy of c at w to be $\alpha(c, w) := \sum_{X \in \mathcal{F}} |c(X) - w(X)|$. More generally, it rules out all of the power measures, $\alpha^z(c, w) = \sum_{X \in \mathcal{F}} |c(X) - w(X)|^z$, except the Brier score, where $z = 2$. And it rules out all of the exponential scores $\pi^z(c, w) = \sum_{X \in \mathcal{F}} z^{|c(X) - w(X)|}$. And it rules out many many more. Thus, the distance between CCD and ETP^+ is substantial. To reach the latter from the former, we must rule out a large number of candidate inaccuracy measures that seem initially plausible — or, at least, as plausible initially as those we end up endorsing. Of course, we offer arguments in favour of ruling out all but the strictly proper inaccuracy measures — in Chapter 4 of *Accuracy and the Laws of Credence*, but also see (Joyce, 2009) and (D'Agostino & Sinigaglia, 2010) — but they are often complex, involved arguments, and this again testifies to how far ETP$^+$ lies from CCD.

Let us now turn to Cariani's second test for dialectical circularity. Like the first, it is based on a reason an interlocutor might have for rejecting a premise of the epistemic utility argument for ETP$^+$. In this case, the reason is that accepting those premises requires her to give up something to which she is strongly committed. Cariani points out that, for some agents, the verdicts of CCD conflict with the verdicts of a more fundamental and plausible principle of decision theory, namely, the principle that demands that we maximise our subjective expected utility — that is, we pick an option with maximal expected utility by the lights of our own credence function. Call this principle EU.

Recall Maya from above. She knows that the chance of H is either between 60% and 70% or between 80% and 90%. Let us suppose that, in violation of ETP$^+$, Maya assigns a credence of 0.25 to H. Now she is offered a bet that the coin will lands heads — £0 if it lands tails, £1 if it lands heads, and the bet costs 50p. CCD demands that she take the bet — every epistemically possible chance function assigns it positive expected utility. EU, in contrast, demands that she doesn't — Maya's own credences assign it negative expected utility. So CCD and EU conflict for her. Cariani suggests that this makes the argument dialectically circular — the interlocutor might very reasonably say that she prefers to preserve EU than to adopt CCD.

The beginning of my response is this: EU and CCD conflict only for agents who violate ETP$^+$. Indeed, if you satisfy ETP$^+$, CCD follows from EU. So, if my interlocutor were to accept CCD, she could then derive ETP$^+$ via the epistemic utility argument, and then she could adopt EU, safe in the knowledge that it will never conflict with CCD *for her*. So, in that sense, I am not asking her to abandon EU. If she accepts the premises and thus the conclusion of my argument, she can retain her allegiance to EU for herself.

But perhaps I am asking my interlocutor to abandon EU in its full generality, and perhaps that might be bad enough? I think not. CCD and EU both state requirements of rationality; they both give necessary conditions for an agent to be rational. That is, they both specify a certain subset of the available options and declare those amongst the irrational options. In Maya's case, those two subsets are disjoint — every option is ruled irrational by CCD or by EU. Thus, there is no option that is not irrational for Maya. But that should come as no surprise to us — after all, Maya is herself irrational; her credences are not appropriate responses to her total evidence. The conflict between CCD and EU in Maya's case, therefore, is simply a symptom of the irrationality of her credences. It gives us no reason to abandon EU.

A similar phenomenon occurs for agents whose credence functions are not probabilistic. Let R be the proposition that it is raining, and H the proposition that the coin in my hand will lands heads the next time it is tossed. Suppose my credences are as follows:

HR	$H\overline{R}$	$\overline{H}R$	$\overline{HR}$	H	$\overline{H}$
0.25	0.25	0.25	0.25	0.9	0.1

That is, they are non-probabilistic, because $c(H) \neq c(HR) + c(H\overline{R})$ and $c(\overline{H}) \neq c(\overline{H}R) + c(\overline{HR})$. I am then offered a bet on H: £0 if $\overline{H}$; £1 if H; and the bet costs 70p. If I maximize my expected utility relative to my credences in the more fine-grained partition (HR, $H\overline{R}$, $\overline{H}R$, $\overline{HR}$), I won't buy the bet — it has expected value of $(0.25 \times 70\text{p}) + (0.25 \times 70\text{p}) +$

$(0.25 \times -30\text{p}) + (0.25 \times -30\text{p}) = -20\text{p}$. But if I maximize my expected utility relative to my credences in the more coarse-grained partition $(H, \overline{H})$, I will — it has expected value of $(0.9 \times 70\text{p}) + (0.1 \times -30\text{p}) = 60\text{p}$. So EU itself can give conflicting advice to agents with non-probabilistic credence functions. Of course, my interlocutor might respond that EU should be restricted to probabilistic agents. But then I see no reason not to restrict it further to agents that also satisfy ETP^+.

Cariani's critique of the epistemic utility argument for ETP^+ helps greatly to clarify the structure of that argument and the responses one might give to an interlocutor who is sceptical of its conclusion. However, I submit that the argument survives the objections. Even if ETP^+ is not best described as a chance-credence principle, it is certainly a principle of credal rationality, and the epistemic utility argument given in *Accuracy and the Laws of Credence* provides strong non-circular support for it.

5 Reply to Horowitz

Link to Horowitz contribution

At first sight, Sophie Horowitz's excellent paper seems to be an essay of two halves. The first concerns the extent to which an individual agent can obey all of the laws of credence laid out in *Accuracy and the Laws of Credence*, and the relationship between the various decision-theoretic principles to which I appeal when I give my epistemic utility arguments in favour of those laws. The second half of the paper concerns the epistemic utility arguments for the Principle of Indifference discussed in Chapter 12 more particularly. However, as we will see, the two halves are related in an unexpected way.

Let's start with the first. Recall that each epistemic utility argument in *Accuracy and the Laws of Credence* consists of (a) an account of epistemic value and (b) a principle of decision theory. As Horowitz notes, there are many principles of decision theory to which we appeal at different points throughout the book; and it isn't clear which principles are compatible with each other. Indeed, I note at the end of chapter 13 that two of the central principles conflict; and this raises the question whether others may as well. As Horowitz puts it, the book might seem to hold out the promise of a unified tour through a range of mutually consistent credal laws. But in fact it presents a range of bespoke packages of such principles, allowing you to choose your own credal adventure.

The central decision-theoretic principles at play in the book are four. We follow the practice of the book and present these as general decision principles that evaluate decisions over a range of options. These options could be actions the agent might perform, in which case the utility would be all-things-considered subjective utility; or they could be credal states she might adopt, in which case the utility would be epistemic utility; or they could be scientific theories she might accept, in which case the utility would be some combination of the epistemic and practical value of accepting a theory. The characterizations that follow here are rough — they are given with greater precision in the book.

- **Dominance**: if o^* has greater *utility* than o at *every possible world*, then o is irrational.
- **Chance Dominance**: if o^* has greater *expected utility* than o by the lights of *every possible objective chance function*, then o is irrational.

- **Maximize Subjective Expected Utility**: if o^* has greater *expected utility* than o by the lights of *the agent's own credence function*, then o is irrational.
- **Maximin**, **Maximax**, **Hurwicz criterion**$_\lambda$ are risk-sensitive principles.
  - **Maximin**: if o^* has greater utility in its worst-case scenario than o has in its worst-case scenario, then o is irrational.
  - **Maximax**: if o^* has greater utility in its best-case scenario than o has in its best-case scenario, then o is irrational.
  - **Hurwicz criterion**$_\lambda$ lies somewhere inbetween. Let the *Hurwicz compromise*$_\lambda$ of an option be the result of weighting its best-case utility by λ and its worst-case utility by $1 - \lambda$. **Hurwicz criterion**$_\lambda$: if o^* has greater Hurwicz compromise$_\lambda$ than o, then o is irrational.
    Maximin and **Maximax** are special cases of **Hurwicz criterion**$_\lambda$:
    Hurwicz criterion$_1$ = **Maximax**.
    Hurwicz criterion$_0$ = **Maximin**.

In addition to these, there are the hybrid principles.

- C-**Maximin**. Where **Maximin** asks us to maximise worst-case utility, C-**Maximin** asks us to maximise objective expected worst-case utility. **Maximin** judges each option by the utility it has in the world where it has lowest utility, and seeks the option that is best in those terms. C-**Maximin** judges each option by the expected utility it has by the lights of the objective chance function that assigns it lowest expected utility, and seeks the option that is best in those terms.
- C-**Maximax** judges each option by the expected utility it has by the lights of the objective chance function that assigns it highest expected utility, and seeks the option that is best in those terms.
- C-**Hurwicz criterion**$_\lambda$ judges each option by a weighted average of its worst-case expected utility and its best-case expected utility, and seeks the option that is best in those terms.

These are the decision-theoretic principles. Together with our account of epistemic value, here's what each entails:

- **Dominance** gives Probabilism (Theorem 4.3.4) and Conditionalization (Theorem 14.3.1).[5]
- **Chance Dominance** gives the Principal Principle (or one of its variants) (Theorem 10.1.1).[6]
- **Maximise Subjective Expected Utility** gives Conditionalization (Theorem 14.1.1).[7]
- **Maximin** gives the Principle of Indifference; **Maximax** demands that an agent be maximally opinionated; **Hurwicz criterion** gives an appropriate compromise between indifference and opinionation (Theorem 13.1.1).[8]

[5]Original results in (Predd et al., 2009) and (Briggs & Pettigrew, ms).
[6]Original result in (Pettigrew, 2013).
[7]Original result in (Greaves & Wallace, 2006).
[8]Original results in (Pettigrew, 2014) and (Pettigrew, 2016).

- C**-Maximin** gives the Principal Principle and, in some cases, a natural version of the Principle of Indifference over chance hypotheses (Theorem 13.3.1).

But which of these decision-theoretic principles are compatible with each other, and which conflict? In Horowitz's metaphor, which adventures can we choose? Here is an attempt at a summary:

- **Chance Dominance** is stronger than **Dominance**. That is, everything that is ruled irrational by the latter is ruled irrational by the former; and, unless the chance functions are all trivial, there will be something ruled irrational by the former that is not ruled irrational by the latter. Neither of these principles appeals to the agent's credences.
- **Maximize Subjective Expected Utility** is also stronger than **Dominance**. This principle does appeal to the agent's credences.
- **Maximize Subjective Expected Utility** is also stronger than **Chance Dominance**, if the agent's credences satisfy the Principal Principle.
- **Maximin**, **Maximax**, and **Hurwicz criterion** are all stronger than **Dominance**. None of these appeal to credences.
- **Maximin**, **Maximax**, and **Hurwicz criterion** are all incompatible with **Chance Dominance**. That is, there are some situations in which they jointly rule out all credence functions as irrational.
- **Maximin**, **Maximax**, and **Hurwicz criterion** are all incompatible with **Maximize Subjective Expected Utility**. That is, there are some situations in which they jointly rule out all credence functions as irrational.
- C**-Maximin**, C**-Maximax**, and C**-Hurwicz criterion** are all stronger than **Chance Dominance**.
- C**-Maximin**, C**-Maximax**, and C**-Hurwicz criterion** are all incompatible with **Maximize Subjective Expected Utility**.

So, it seems that we might have a number of package tours available for the epistemically discerning traveller. They roughly divide as follows: in the first grouping are those in which you are risk-neutral throughout your epistemic life; the second grouping consists of those that require risk-sensitivity at the beginning of your epistemic life, and risk-neutrality thereafter; and those in the third grouping require a certain sort of risk-sensitivity throughout.

- Group 1 — always risk-neutral.
  - Package 1a: You adhere to **Dominance**, **Chance Dominance**, and **Maximize Subjective Expected Utility** at all times throughout your epistemic life. So you obey Probabilism and the Principal Principle; and you plan to update by Conditionalization (because that's the update rule that minimizes expected inaccuracy, but also because of the accuracy dominance argument in its favour).
- Group 2 — risk-sensitive as a superbaby; risk-neutral thereafter.

- Package 2a: You adhere to **Dominance** and **Maximin** at the beginning of your epistemic life; and thereafter you adhere to **Maximize Subjective Expected Utility**. So you obey Probabilism and the Principle of Indifference; and you plan to update by Conditionalization (because that's the update rule that minimizes expected inaccuracy, but also because of the accuracy dominance argument in its favour).
- Package 2b: You adhere to **Dominance** and some non-extremal version of **Hurwicz criterion** at the beginning of your epistemic life; and thereafter you adhere to **Maximize Subjective Expected Utility**. So you obey Probabilism and you adopt an initial credence function that effects some compromise between indifference and maximal opinionation; and you plan to update by Conditionalization (because that's the update rule that minimizes expected inaccuracy, but also because of the accuracy dominance argument in its favour).

- Group 3 — always risk-sensitive.
  - Package 3a: You adhere to **Dominance** and **Maximin** throughout your epistemic life. So you obey Probabilism and the Principle of Indifference; and you plan to update by Conditionalization (because that's the update rule that maximizes worst-case accuracy, but also because of the accuracy dominance argument for that update rule). But what's more, providing you formulate **Maximin** in the correct way, it leads you actually to update by Conditionalization, as well as planning to. **Maximin** says that you should pick the option whose worst-case scenario is best. But this is clearly relative to a set of possible worlds — it tells you to pick the option whose worst-case scenario *amongst those worlds* is best. If we pick the entire set of possible worlds, **Maximin** requires equal credence in all worlds. However, if we pick just the set of worlds at which our agent's evidence is true, then **Maximin** requires equal credence in all of those worlds, together with credence 0 in all other worlds. The correct version of **Maximin** is relativised to the set of worlds at which the agent's evidence is true. Thus, for an agent with evidence E, **Dominance** and the correct version of **Maximin** require that she adopt credence function $c^\dagger_E$, where $c^\dagger_E(X)$ is just the proportion of E-worlds at which X is true. But it is clear that, if a later piece of evidence E' is stronger evidence than the agent's current evidence E, then $c^\dagger_{E'}(-) = c^\dagger_E(-|E')$. That is, the agent updates by Conditionalization upon receipt of the stronger evidence. So an agent who obeys this evidence-relative version of **Maximin** will in fact adopt the credences that she would if she were updating by Conditionalization on an initial uniform distribution over the whole set of possible worlds — she won't just plan to conditionalize; she'll actually do it.
  - Package 3b: You adhere to **Dominance** and some non-extremal version of **Hurwicz criterion** throughout your epistemic life. So you obey Probabilism and you adopt an initial credence function that effects some compromise between indifference and maximal opinionation; and you plan to update by Conditionalization (because of the accuracy dominance argument in its favour). However, we then run into a problem. In Part IV of *Accuracy and the Laws of Credence*, I describe three different arguments that agents should plan to update by Conditionalization. It is the first and third that concern us here. In the first, we use our prior credences —

those we have before we learn the new evidence — to assess the expected inaccuracy of possible update rules; we note that the rule that minimizes that expected inaccuracy is the Conditionalization rule for that prior. In the third argument, we do not use our priors to assess update rules; rather, we assess pairs consisting of priors and update rules together; we note that such a pair is only undominated if the update rule is the Conditionalization rule for the priors. Thus, these two quite different ways of assessing update rules make the same demand. However, it turns out that this is not always the case. If we replace **Maximize Subjective Expected Utility** with our non-extremal version of the **Hurwicz criterion** in the first argument, it does not recommend Conditionalization, even if the priors are those that satisfied that version of the **Hurwicz criterion**. But if we replace **Dominance** with the non-extremal version of the **Hurwicz criterion** in the third argument, it does recommend Conditionalization, because **Hurwicz criterion** is stronger than **Dominance**. Thus, adopting the **Hurwicz criterion** throughout your epistemic life creates a dilemma — which of the two ways of assessing update rules is correct? This isn't necessarily the sort of incompatibility that Horowitz has in mind in her essay; but it is revealed when we really attend to the relationships between the different laws of credence and principles of decision theory in the way that Horowitz demands.[9]

– Package 3c: You adhere to **Dominance**, **Chance Dominance**, and C**-Maximin** throughout your epistemic life. So you obey Probabilism and the Principal Principle; and in some circumstances, where your evidence about the chances is symmetric in some sense, you obey a version of the Principle of Indifference that demands equal credence over the various chance hypotheses; what's more, you plan to update by Conditionalization (because of the accuracy dominance argument in its favour). However, here again we run into difficulties. If the correct inaccuracy measure is the local log score, then everything is fine: you will not just plan to update by Conditionalization; you will in fact also update by conditionalization, since that is the effect of adhering to C**-Maximin** throughout your life (Williams, 1980; Seidenfeld, 1986; Grünwald & Dawid, 2004). But if the correct inaccuracy measure is the Brier score (as I argue in chapter 4 it should be), then, while we will plan to update by Conditionalization, we will in fact update by a different rule (Grünwald & Dawid, 2004, Example 7.1). Thus, as with Package 3b, we face an epistemic dilemma.

I'll leave this survey here. Hopefully it goes some way to providing the sort of catalogue of compatibilities and incompatibilities that Horowitz rightly requests. And it fully vindicates her contention that *Accuracy and the Laws of Credence* does not provide a single unified creed for a virtuous epistemic life, but rather something of a motley of distinct denominations between which an agent must pick. Hopefully the survey I have just outlined allows such an agent to see exactly what they might sign up to.

Let's move on now to the second half of Horowitz's paper. Horowitz is here concerned with my motivation for **Maximin**, and my argument for the Principle of Indifference, which

[9]Note that recent unpublished work by Catrin Campbell-Moore and Bernhard Salow identifies a conflict between Conditionalization and another suite of risk-sensitive decision principles, namely, the various risk-sensitive principles that arise when we pick a non-trivial risk function in the framework of Lara Buchak's risk-weighted expected utility theory (Buchak, 2013).

is built upon it. She identifies three assumptions of the argument: (1) principles that govern our credal state in the absence of evidence can be determined by asking what an agent should do who has no credences as well as no evidence — agents that Lewis dubbed *superbabies*; (2) such agents may permissibly adhere to a decision principle in that initial credence-less state that is different from the principle they later use; (3) **Maximin** is a permissible decision principle in that initial credence-less state. Let's look at these in turn.

Contra (1), Horowitz contends that it is hard to imagine an agent devoid of beliefs. Perhaps. But even if that were true, I think it does not undermine my appeal to such an agent. As I emphasise in the Introduction, *Accuracy and the Laws of Credence* seeks to justify laws of rationality, not norms. That is, the negative judgments made in the book against agents who violate Probabilism or the Principal Principle, or who plan to update other than by Conditionalization, are evaluative judgments, not normative ones — they declare the agent faulty, but not necessarily blameworthy. The reason is that I do not subscribe to doxastic voluntarism — I do not think an agent chooses her beliefs or her credences. Rather, she has those credences, and we can evaluate them using the evaluative versions of the decision principles that I appeal to in the book. Sometimes, when we evaluate a particular doxastic state of an agent, there is a credence function that is relevant to that evaluation not just because it is the credence function being evaluated. For instance, when we evaluate the updating rule an agent plans to use, we appeal to the agent's prior credence function in order to calculate the subjective expected inaccuracy of that rule, which we might seek to minimize. But, more often, there is no such credence function. And in that case, we appeal to a decision principle that does not require any credence function as an input — a decision principle like **Dominance**, **Chance Dominance**, **Maximin**, **Maximax**, or the various versions of **Hurwicz criterion**$_\lambda$. In this case, we might talk as if we have an agent with no credences who is choosing which credence function to adopt using that decision principle. But what we really have is an agent who has her initial credence function (whether chosen or not), and whom we then evaluate using the principle. Thus, in the case of the argument for the Principle of Indifference, we do not (typically) have an agent with no credences and no evidence who is choosing her initial credence function on the basis of **Maximin**; rather, we have an agent with no evidence who has her initial credence function (whether chosen or not), and whom we then evaluate using **Maximin**.

Let's turn to Horowitz's concern about assumption (2). I agree with Horowitz that it seems unmotivated to use **Maximin** at the first moment of your epistemic life, and then some other, risk-neutral rule thereafter. And I agree that the Rawls analogy goes little way to justifying it. One notable disanalogy is that, in the Rawlsian framework, we use Maximin to choose long-lasting things, such as the structure of society, the nature of its institutions, and so on. That is, we choose features of it that will be very hard, perhaps impossible, to change once we have emerged from behind the veil of ignorance. And that might give us reason to be risk-averse in making those decisions. That is not the case here. We do not choose features of our credences that we will be bound to preserve throughout our epistemic lives — at any point, we can simply switch them to something else. So I would prefer a picture in which we do not posit such a shift in attitudes to epistemic risk between the initial point in an agent's epistemic life and the other points. Package 3a seems to offer this, at least for those to whom **Maximin** appeals.

And that brings us to Horowitz's final worry, which concerns assumption (3): Is **Maximin** permissible? The problem is that **Maximin** is the most extreme form of risk aversion — it counsels us not just to give greater weight to the worst-case scenario, but to give all

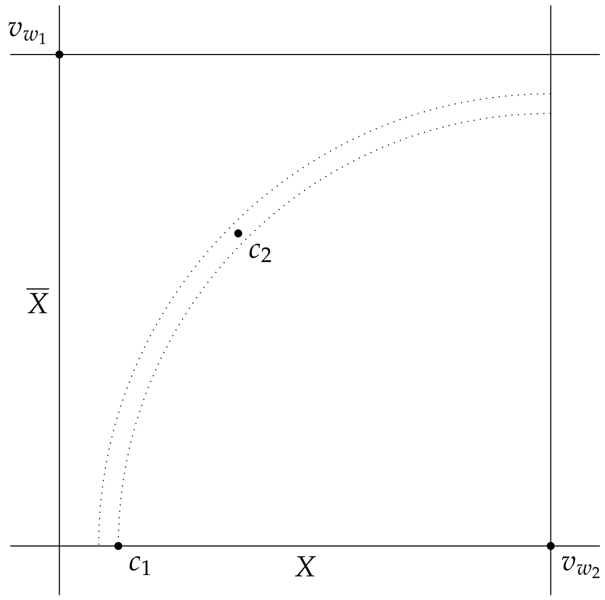

Figure 1: We represent credence functions defined on X and $\overline{X}$ as points in the unit square. v_{w_1} is the omniscient credence function if $\overline{X}$ is true; v_{w_2} is the omniscient credence function if X is false. The right-hand dotted line represents all the credence functions that are as accurate at world w_2 as c_1 is. In between the two dotted lines lie all the credence functions that are only slightly less accurate than c_1 at w_2. As we can see, none of them is dramatically more accurate at w_1. c_2 is one of those for which there is greatest discrepancy between the accuracy of c_1 and c_2 at w_1. And it is not so much more accurate.

of our weight to it. In practical decisions, that leads to absurdly risk-averse outcomes. It demands that I refuse a bet on rain tomorrow that costs £1, pays £1 trillion if it rains, and pays £0 if it doesn't. However, such a situation cannot arise in the credal case, as Figure 1 shows. Suppose I have two credal options, c_1 and c_2. And suppose that c_2 is only very slightly worse than c_1 at world w_2. Then c_2 cannot be dramatically better than c_1 at world w_1. And that's what would be required in order to have an analogous situation to the bet on rain described above. So perhaps **Maximin** is not so absurd in the credal case. However, neither is it plausible enough that it should be the only risk-sensitive decision principle that is permitted. Other points along the spectrum given by **Hurwicz criterion**$_\lambda$ should also be permissible. But then we run into Horowitz's second worry. As we saw above, if you adhere to any non-extremal version of **Hurwicz criterion**$_\lambda$ throughout our epistemic life, your judgments about which update rule to plan to use will be inconsistent — evaluating the updating rule only will lead you to endorse something other than Conditionalization; evaluating prior-updating rule pairs, will lead you to endorse Conditionalization.

So, while Horowitz's paper comes in two halves, the response to one draws on the response to the other in a surprising way. My argument for the Principle of Indifference in the book turns on the assumption that agents have a different attitude to epistemic risk at the beginning of their epistemic life and at other points in it. As we saw in the discussion of Package 3a above, we do not need to assume this; the same laws follow if we assume **Maximin** throughout. However, **Maximin** is an extreme principle. We wish to permit other, less extreme risk-sensitive principles. But, when we assume those, we need to restrict them to the first moment in our epistemic lives in order to avoid an irresolvable dilemma. Further work is needed to assess the consequences of this. Horowitz's paper does a great service in drawing out the problem.

References

Briggs, R., & Pettigrew, R. (ms). Conditionalization. Unpublished manuscript.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

Caie, M. (2015). Credence in the Image of Chance. *Philosophy of Science*, *82*(4), 626–648.

Clarke, R. (2013). Belief is Credence One (in Context). *Philosophers' Imprint*, *13*(11), 1–18.

D'Agostino, M., & Sinigaglia, C. (2010). Epistemic Accuracy and Subjective Probability. In M. Suárez, M. Dorato, & M. Rédei (Eds.) *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association*, (pp. 95–105). Springer Netherlands.

de Finetti, B. (1974). *Theory of Probability*, vol. I. New York: John Wiley & Sons.

Greaves, H., & Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, *115*(459), 607–632.

Grünwald, P. D., & Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *The Annals of Statistics*, *32*(4), 1367–1433.

Hall, N. (1994). Correcting the Guide to Objective Chance. *Mind*, *103*, 505–518.

Hall, N. (2004). Two Mistakes About Credence and Chance. *Australasian Journal of Philosophy*, *82*(1), 93 – 111.

Ismael, J. (2008). Raid! Dissolving the Big, Bad Bug. *Noûs*, *42*(2), 292–307.

Ismael, J. (2013). In Defense of IP: A response to Pettigrew. *Noûs*.

Jeffrey, R. C. (1970). Dracula meets Wolfman: Acceptance vs. Partial Belief. In M. Swain (Ed.) *Induction, Acceptance, and Rational Belief*, (pp. 157–185). Reidel.

Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, *65*(4), 575–603.

Joyce, J. M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber, & C. Schmidt-Petri (Eds.) *Degrees of Belief*. Springer.

Kierland, B., & Monton, B. (1999). Minimizing Inaccuracy for Self-Locating Beliefs. *Philosophy and Phenomenological Research*, *70*(2), 384–395.

Lange, M. (1999). Calibration and the Epistemological Role of Bayesian Conditionalization. *The Journal of Philosophy*, *96*(6), 294–324.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence Accumulation in Decision Making: Unifying The "Take the Best" and The "Rational" Models. *Psychonomic Bulletin & Review*, *11*(2), 343–352.

Levinstein, B. A. (2015). With All Due Respect: The Macro-Epistemology of Disagreement. *Philosophers' Imprint*, *15*(3), 1–20.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In R. C. Jeffrey (Ed.) *Studies in Inductive Logic and Probability*, vol. II. Berkeley: University of California Press.

Lewis, D. (1994). Humean Supervenience Debugged. *Mind*, *103*, 473–490.

Moss, S. (2011). Scoring Rules and Epistemic Compromise. *Mind*, *120*(480), 1053–1069.

Pettigrew, R. (2013). A New Epistemic Utility Argument for the Principal Principle. *Episteme*, *10*(1), 19–35.

Pettigrew, R. (2014). Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research*.

Pettigrew, R. (2015). Accuracy and the Credence-Belief Connection. *Philosophers' Imprint*, *15*(16), 1–20.

Pettigrew, R. (2016). Jamesian epistemology formalised: an explication of 'The Will to Believe'. *Episteme*, *13*(3), 253–268.

Predd, J., Seiringer, R., Lieb, E. H., Osherson, D., Poor, V., & Kulkarni, S. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions of Information Theory*, *55*(10), 4786–4792.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading: Addison-Wesley.

Ramsey, F. P. (1931). Truth and Probability. *The Foundations of Mathematics and Other Logical Essays*, (pp. 156–198).

Schoenfield, M. (ta). An Accuracy-Based Approach to Higher-Order Evidence. *Philosophy and Phenomenological Research*.

Seidenfeld, T. (1986). Entropy and Uncertainty. *Philosophy of Science*, *53*(4), 467–491.

Shimony, A. (1988). An Adamite Derivation of the Calculus of Probability. In J. Fetzer (Ed.) *Probability and Causalty*. Reidel.

Sinnott-Armstrong, W. (1999). Begging the Question. *Australasian Journal of Philosophy*, *77*(2), 174–191.

Staffel, J. (2015). Disagreement and Epistemic Utility-Based Compromise. *Journal of Philosophical Logic*.

Thau, M. (1994). Undermining and Admissibility. *Mind*, *103*, 491–504.

Tversky, A. (1972). Elimination by Aspects: A Theory of Choice. *Psychological Review*, *79*(4), 281–299.

van Fraassen, B. C. (1983). Calibration: Frequency Justification for Personal Probability. In R. S. Cohen, & L. Laudan (Eds.) *Physics, Philosophy, and Psychoanalysis*. Dordrecht: Springer.

Weisberg, J. (ta). Belief in Psyontology. *Philosophers' Imprint*.

Williams, P. M. (1980). Bayesian Conditionalization and the Principle of Minimum Information. *The British Journal for the Philosophy of Science*, *31*, 131–144.